Erasmus+

Andrea Favia, Paolo Berizzi

1 Introduction

The Erasmus program - EuRopean Community Action Scheme for the Mobility of University Students - was launched in 1987 and in the first year it brought more than three thousand students abroad.
The program fosters individuals' self-development and, as a consequence, a better cohesion among countries. Also, the student is able to build international networking and predisposition toward working in a different country in the future. The number of people participating in the program is steadily increasing and in the latest annual report dated 2016, Erasmus supported 725 000 mobilities, reaching more than 79 000 organizations and new education policies were launched.
The budget for the mobility for the period 2007-2013 was about EUR 3 billion while for the period 2014-2020, it reached more than EUR 14 billion, showing how important this program is becoming not only for students but for entire Europe. People can choose among different opportunities such as exchanges program, joint degrees, traineeships or do voluntary work.

1.1 Datasets

EU Open Data Portal [1]offers yearly dataset where it is possible to gain a broad and in-depth overview of the Erasmus initiative. Given a specific year, for each anonymized student participating in the program, it is possible to see the original university and the destination, as well as other features such as the study program and personal information.
For our purpose, we first downloaded the annual dataset published for the period 2009-2012 and we combined them together to include in our analysis the temporal dimension. We selected the one above mentioned essentially because they presented the same structure (i.e. same columns attributes) [4] [5] [6] [7].

In addition, since the organizations were encoded by an institution erasmus code, we had to find a datasets [source] to map them to actual geographic positions. Even though we used the up-to-date dataset of erasmus institutions, some of them have not been recognized, leaving about 8 thousand observations without a mapping.

Moreover, we needed an additional mapping for the field of study for each student and therefore we used the dataset found on the Erasmus portal [source] for the additional mapping.

Lastly, we decided to have in our analysis the total number of students for different countries and different year and we found these data on the UN open data platform [source]. In this way, it is possible to meaningfully [9] compare countries with relative magnitude, rather than absolute values.

• *Andrea Favia and Paolo Berizzi are with Eindhoven University of Technology. E-mail: {a.favia, p.berizz}@student.tue.nl.*

1.2 Merging and Data Cleaning

In order to perfectly merge the dataset we had to correct some columns names and the order in which they were presented.

Also, we removed some typos in the spelling of some countries and used, for every datasets, the country codes defined by the *ISO 3166-1 alpha-2* so that the join operations could be done.

The simple data cleaning has been performed via python libraries (i.e. Pandas and Numpy) and the joins / merges among datasets have been realized through Tableau, the software we have used for the entire visualization analysis.
It is possible to see all the features we have used in the resulting dataset in Table 1. There are in total 968 507 observations but, due to the unsuccessful mapping, we have 3 200 missing values. In the Table 2 it is possible to have a better idea of the number of observations grouped by different years. We decided not to use the missing values since it was tricky and difficult to input manually all the institutions not directly mapped and we are still able to use 965 307 observations. Likely, these institutions we could not map are not partnering with Erasmus anymore and therefore we concluded that they are not essential for the analysis.

Table 1. Definition and description of data attributes used during the analysis

DATA ATTRIBUTES	DATA DESCRIPTION
ID	ID of the mobility
GENDER	Gender of student
MOBILITYTYPE	Type of mobility: Study, Placement, Combined
YEAR	Year of the mobility: 2009, 2010, 2011, 2012
COUNTRYOFWORKPLACEMENT	The country where the placement took place
COUNTRYOFHOMEINSTITUTION	Country of the home institution where a student is registered
COUNTRYOFHOSTINSTITUTION	Country of the host institution where a student has studied
LENGTHSTUDYPERIOD	How long (in months) a student has been abroad
ECTSCREDITSWORK	Number of ECTS Study credits anticipated
ECTSCREDITSSTUDY	Number of ECTS for Placement credits anticipated
STUDYGRANT	The amount of grant the student received for study period
PLACEMENTGRANT	The amount of grant the student received for placement period
MAINFIELD [a]	The main field of study the student belongs to at the home institution
DESCRIPTION[a]	The study program of the student at the home institution
HOME_ORGANIZATION_NAME	The name of Home Institution where the student is registered
HOME_CITY	The name of the city in which Home institution is located
HOST_ORGANIZATION_NAME	The name of Host Institution where the student is registered
HOST_CITY	The name of the city in which Host institution is located
TOT_POPULATION	The total student population for a particular tuple (COUNTRYOFHOMEINSTITUTION, YEAR)

[a] Example of MAINFIELD is Health and Welfare
[b] Example of DESCRIPTION is Medicine

2 Problem Description and Task Analysis

Having a visualization can easily allow a future or potential Erasmus student to gather information about the universities, the environment in which he might be surrounded and the entire

Table 2. Number of observations used and missing values for different years

YEAR	TOTAL OBSERVATIONS USED	MISSING VALUES
2009	213 030	1 405
2010	230 822	1 244
2011	253 137	500
2012	268 318	51

ecosystem. Also, an in-depth analysis can allow Erasmus employees to grasp discrepancies between different countries or universities and therefore make new policies and try to focus more on certain characteristics. Are money granted to students evenly distributed among different countries? Are there universities that are preferred more? If so, where they are located and is the field of study an important factor in order to chose an university as destination?

In order to answer most of the answers above mentioned and grasp other information, we define and describe three distinct tasks by using the taxonomy presented in the work of Hans-Jorz et al. [8]. In this case, each task will be defined in a 5-space dimension that relates to the WHY, WHAT, WHO, WHERE, WHEN of our visualization. Hence, each task will be summarized with the 5-tuple $(goal, mean, characteristic, target, cardinality)$. A detailed description of the taxonomy and the words used to describe each section can be found on the original paper [8]. In more details, our work presents workflows of tasks, since some individual tasks follow other individual tasks. We must notice that out primary target for the visualization is an audience that does not have expert knowledge about the Erasmus organization, such as a curious student that would love to get more in-depth information. As a result, we mostly do *exploratory* analysis but it can directly be interpreted as a *confirmatory* analysis for an expert audience.

2.1 In-going and Out-going students flow

Our first question was to understand the in-going and out-going flows of students for different countries. Specifically, given a particular place, we wanted to get an overview of the number of students going to or coming from that particular one. Furthermore, we wanted to allow the user to get an idea about which students (grouped by Field of study and Gender) constitute each particular flow.

$$\begin{gathered}(exploratory|presentation,\ search|navigate,\ frequency,\\ attr(\texttt{COUNTRYOFHOSTINSTITUTION})|\\ attr(\texttt{COUNTRYOFHOMEINSTITUTION}),\ all)\\ \Downarrow\\ (exploratory,\ summarize,\ frequency,\\ attr(\texttt{MAINFIELD})|attr(GENDER),\ single)\end{gathered} \quad (1)$$

2.2 In-depth EDA for students grouped by country, city, and institutions

Given a country, city or institution it is possible to see the number of outgoing or ingoing students. Ideally, the user might first think about a country and then narrow down his research to find a particular university. On the other hand, the user may decide to pick a particular city or directly an institution, maybe because he already knows the destination or driven by a curiosity about his own university. As a subtask, we allow the user to investigate some data about a particular university. Namely, the user can navigate to see summaries of the most preferred universities (i.e. university that host highest number of Erasmus students) and the most common field of study. In addition, he would be able to use each of these data as a filter to get a deeper level of detail. Hence, we defined our workflow of tasks as following.

$$\begin{gathered}(epxloratory,\ search|filter|summarize,\ frequency,\\ attrib(outgoingstudents)|attrib(ingoingstudents)|\\ attib(\texttt{HOME\_CITY})|attr(\texttt{HOME\_ORGANIZATION\_NAME}),\ all)\\ \Downarrow\\ (exploratory,\ search|summarize,\\ frequency|trends,\ attr(\texttt{TOT\_ETCS})|attr(\texttt{HOST\_ORGANIZATION\_NAME})|\\ attr(\texttt{MAIN\_FIELD}),\ multiple)\\ \Downarrow\\ (presentation,\ summarize,\ frequency,\ attr(\texttt{GENDER}),\ single)\end{gathered} \quad (2)$$

2.3 Comparison of Erasmus students characteristics among different countries

Finally, we considered essential to allow the user to compare data among different countries. Hence, our task addresses the problem to identify dissimilarities or similarities in the grant received, in the length of the period spent abroad, and also about the gender distributions. We must point out that in order to have a meaningful comparison between different countries [9], we had to use relative scales rather than absolute ones. This justifies the use of the additional dataset on the total population of students for different countries and years. In the end, our final tasks overflow is the following:

$$\begin{gathered}(exploratory,\ compare|search|navigate, distribution\\ |frequency|trend,\\ attr(\texttt{LENGTH\_STUDY\_PERIOD})|attr(\texttt{LENGTH\_WORK\_PERIOD})|\\ attr(\texttt{STUDY\_GRANT})|\\ attr(\texttt{PLACEMENT\_GRANT})|attr(tot\_students), all)\\ \Downarrow\\ (exploratory,\ summarize,\ frequency,\ attr(tot\_grant), single).\end{gathered} \quad (3)$$

3 Visualization Design

To offer a clear and easy-to-use solution that could at the same time provide non-trivial insights through interaction we decided to use Tableau. That choice gave us the possibility to make proper use of the details offered by the dataset and realize a powerful interactive application without incurring in software too complex to understand for the user. For each of the tasks previously defined we have created a different dashboard composed of a set of integrated visualization. In that way we could clearly answer the problem defined by the tasks as well as give the context needed to fully understand each visualization.

3.1 Task 1

The first task allows the user to discover the variety of flows of students moving from one country to another and understand the differences among these movements. Specifically, it is possible to *filter* for a particular year and mobility type, enabling a general comparison over the time and between students and workers. The

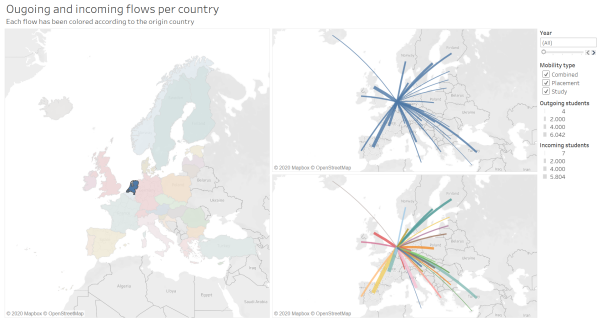

Fig. 1. Dashboard presenting the outgoing and incoming students per country, after having selected *The Netherlands* as a country

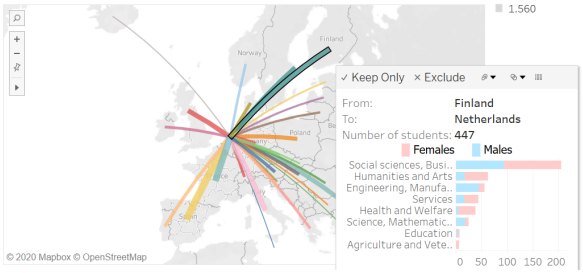

Fig. 2. Detail of incoming students for The Netherlands and tooltip for the students coming from Finland

entire dashboard is shown in Figure1 where we imagine the user want to analyze The Netherlands.

The user can easily *lookup* a particular country and subsequently *browse* among the variety of flows related to that country or *lookup* for a specific flux he is interested in. In the main map each country is identified by a color and once a country is *selected*, two more origin-destination maps pop up on the side of the screen presenting the outgoing and incoming students. Each flow is identified by a line that link the two countries: the color of the line encode the origin country according to the color defined in the main map, whereas the number of students belonging to a flow is encoded by the width of the line. The coherent use of the color for the origin-destination lines facilitate the user to *navigate* among different pairs of country, while the thickness of the lines is essential to compare different flows in each map.

The user can perform a more detailed analysis by *hover* on a particular flow: a tooltip shows the number of students belonging to the selected flow and present a bar-chart that groups the student by main field and, at the same time, encode the gender for each bar by using colors. It is thus possible to *explore* the different fields of study that characterize each flow an thus *summarize* the main motivations that support the students in their choice of going abroad. A detail of the map of the students that come in The Netherlands is shown in Figure2 where the tooltip for students from Finland is shown.

As a design technique we firstly evaluated the use of a flow map as suggested in [9]. Despite the optimal visualization that can be generated when this technique is used for a few flows, we find it not applicable in our case due to the high number of flows we aimed to present. To overcame this problem, we decided to use an origin-destination flow map and, following the guideline presented in the study of Jenny et al. [10], we chose slightly curved lines since they are more readable by users.

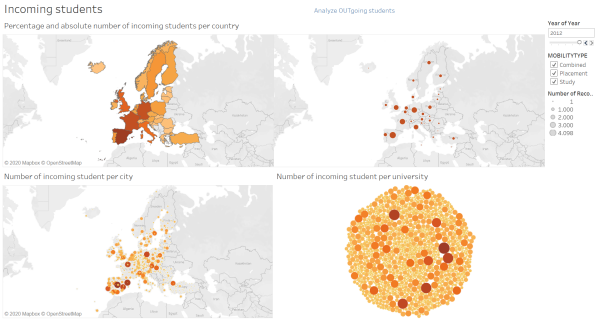

Fig. 3. Dashboard presenting incoming students per country, city and university

When no country has been selected yet, the origin-destination map would simultaneously show all flows in the Europe. This solution would have made it impossible to read the maps as almost all combinations between different countries are present. Instead, in our application, we decided to firstly present a map of Europe with each country identified by a color and, secondly, show the outgoing and incoming flows for the selected country exclusively when the user chooses a particular country in two more tabs that appear on the side of the screen.

We also evaluate the possibility to use Sankey [11] diagrams to represent the flows of students around Europe. Indeed, we found this technique somewhat misleading due to the absence of the geographical coordinate that actually help the user to *lookup* for a particular flow instead of having to locate it as it would be the case in a Sankey diagram.

Regarding filters we use a slider for the years to highlight the progression over time, but we also leave the possibility combine all the years together to have an overview on the whole period of the analysis. For the mobility types we use a multiple value list of students, placement and combined so that the user can chooses any possible combination and draw different conclusions. We use and report two different range for the thickness of the lines regarding outgoing and incoming students as they can differ and doing so ensures a more precise visualization. On the contrary, we did not report the legend that links colors and countries because we think the mapping of these features is already explained by the main map that report all the country with related colours and is still visible even when a particular country is selected.

3.2 Task 2

The second task allows the user to understand how many people come from or are assigned to a particular country, city or university. Furthermore, it is possible to have more information about a specific university, showing some statistics that are going to be defined later.

In a unique view, as the reader can see in 3, the user is able to *filter* for a particular year or mobility type and *lookup* a particular country or city. At the same time the user can *browse* the various maps shown to find the most frequent locations.

Firstly, a choropleth map encodes the percentage of Erasmus students per country with a unique color palette [2]. In particular, the user can *explore* the map to find the highest percentage or *lookup* a particular country. In this case, we had to use relative values (for a specific country) so that from our map is possible to derive significant similarities or discrepancies.

At the same time, we wanted to give the user a concrete idea in term of total number of students for a particular country and

Fig. 4. Zoom of the HostDashboard where *Amsterdam* has been selected

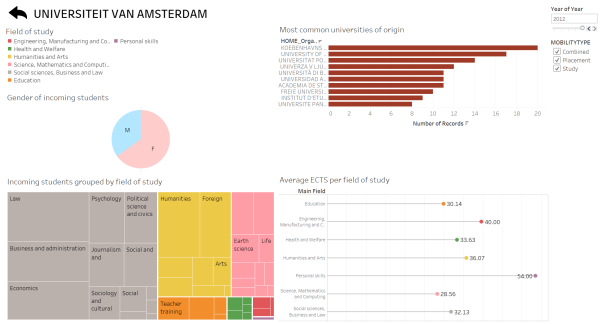

Fig. 5. Dashboard with details for incoming students in *Universiteit Van Amsterdam*

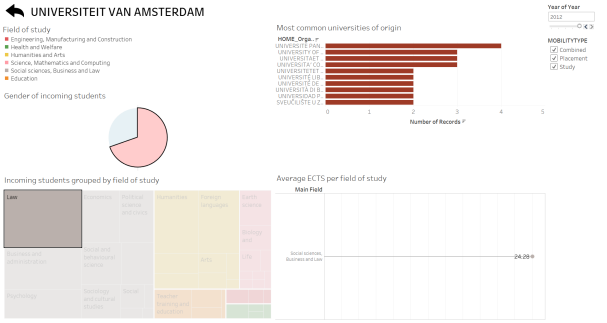

Fig. 6. Dashboard for incoming students in *Universiteit Van Amsterdam* filter on Female and Law

city. Hence, we encoded total number of students as circle in a in two additional symbol maps, both for countries and cities. In this case the user can *explore* the maps to identify the countries and cities with the highest number of outgoing or incoming students.

In order to *identify* the universities with the highest number of Erasmus students (also due to the size of the institution itself), on the bottom-right corner we show a packed bubbles chart that encodes the number of outgoing or incoming student for a particular university. We must point out that our goal was not to compare the total number of students for different universities, but to give a qualitative and general idea of the largest universities. It can be argued that a simple bar chart could convey the same information and allowing a better comparison among different universities. Nonetheless, trough a packed bubbles chart we can give the user a direct and clear overview by encoding both area and color saturation. In Figure 4 a detail of the dashboard is shown after the Netherlands have been selected as a country and Amsterdam has been selected as a city.

Once identified a particular university, the user can click a banner to pop up another view that presents more details on that particular university. Specifically, the dashboard present four more visualizations and is shown in Figure5

A Treemap is used to illustrate information about the main field of study and the precise university program. Since these data present a hierarchy, our choice fell on either a multi-level pie chart or a tree map. We preferred the latter since it is not circular and therefore it is easier to read. The area of the single box encodes the total number of students belonging to a particular program (defined by attribute `description`) and the color encodes the main field of study. Also in this case, a direct comparison between different programs is not easy but our choice highlights patterns in the main field, showing which are the program most frequented.

To understand in which universities people go the most, we used a horizontal bar chart that show the 10 most preferred universities. The choice of the horizontal chart is driven by the necessity to show complete labels, avoiding to rotate them in the vertical version of the bar chart. Nevertheless, the user can *change* the order of the visualization on his own preferences.

Additionally, we investigated the average number of credits (ECTS) acquired by the student for different fields, to understand whether the mobility characteristic depends on the subject area. We convey these information by using a horizontal lollipop chart. As already pointed out beforehand, the horizontal chart allow the user to read quite long labels and the lollipop version is used to avoid bars of the same height to be close to each other and therefore elude the Moiré effect [13].

As final detail, we used a simple pie chart to convey the information about the gender for a particular university.

We have to point out that it is possible to *select* different elements in the visualization and, more important, each of the element present in the final dashboard works as a filter. For example, a user could decide to exclude males and to focus only on people enrolled in Law: in this case he would easily understand how many girls enrolled in Law are present in the Erasmus program for that particular institution and a particular year. Also, it would be possible to find the most common destination and the specific number of ECTS gained for that particular group of interest as it is shown in figure

3.3 Task 3

The last task allows the user to compare different countries to get some insights about the study grant obtained by a student, the *trend* of the number of students throughout the 4-year period considered. Also, it shows the *distribution* of the length of the mobility compared to the average length study period for all the countries in the Erasmus program. In this case, as in the other tasks, it is possible to filter for the mobility type and the year. It is possible to see the dashboard in Figure 7.

On the dashboard, it is possible to identify particular countries and selecting them by clicking on them or by dragging the map. As soon as the countries have been selected, three charts show the information above mentioned. Firstly, in the top-right corner a boxplot encodes the length of the mobility period and therefore it is possible to have a clear idea of the *distribution* of that variable. The user can intuitively and directly understand how many months people of each country spend abroad and compare different countries. Moreover, a red line shows the average length of period for all the countries and each boxplot has its own black line that displays, on hover, the average for the country. The box plot choice is driven by the need to summarize large numbers of observations and the ability to provide a 5-number summaries: minimum and maximum values, lower and upper quartiles, and the median - essential summaries to get an idea of the shape of the

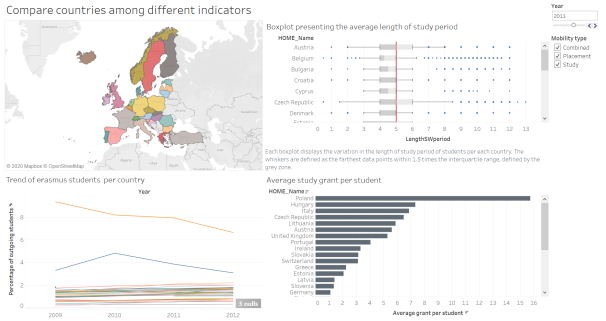

Fig. 7. Dashboard for compare indicators of different countries

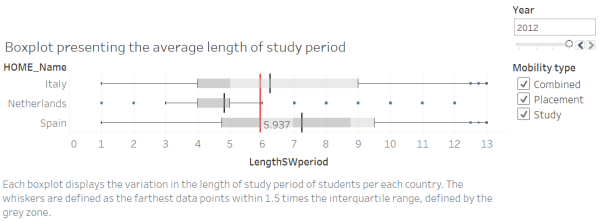

Fig. 8. Box-plot to compare the length period of outgoing students from Netherlands, Spain and Italy

distribution. We must notice that the possible outliers identified cannot be actually considered extreme observations since the underlining distributions are not close to a normal distribution. A representation of the box plot is shown in Figure8.

Secondly, in the bottom left side, a line chart shows the trends of the number of students participating to the mobility program throughout the entire period. Selected a particular combination of countries (each of them encoded with a particular color), a line for each country with the respective color appears. By hovering over a line, an area chart pops up to displays the frequency of students grouped by different fields of study. In this case, the user can easily understand which is the most common department of study. We could have used a bar chart for both the sub-tasks but since the user might want to compare a lot of countries at the same time, the visualization would have been confusing and not readable. Additionally, the area chart is chosen because we want to show the magnitude, in absolute values, of the trend and the name of each field of study.

Lastly, in the bottom right side, for the selected countries a horizontal bar chart shows the average grant per person as the sum of all the subsidies (`STUDYGRANT+PLACEMENTGRANT`). The bars are sorted in descending order so the user can directly *browse* the country with the highest average grant and this also facilitate the comparisons between different countries. In addition to that, once the user has identified a desired country by hovering on in, a vertical histogram shows the grant obtained by the students divided by gender. We did so because the average can sometimes mislead the interpretation and strictly depends on the length of the period, thus we think that showing the complete distribution of grant makes the understanding much more complete.

4 Use Cases

As a general hypothesis we assume that our user is curious about the Erasmus program and would like to get more information so he can better understand the dynamics: he could be a student that

Fig. 9. Detailed of the analyses discussed in use case 4.1

aim to participates to the program and has to choose a particular university, a journalist that is investigating the project to write a report or any other person that is interested in it.

4.1 A student from Italy

As a first use case we imagine the user to be a student from Italy, in particular from the University of Bari, where she study Computer Science. She would like to go abroad for one semester, but she is still far from the final choice of the foreign university so she want firstly a general overview of the possibilities and secondly some more insights about some universities.

Given that, she starts by looking at the *Destination dashboard* that presents the different flows that characterize a country. She filter on all year as she is not interested in any particular trends over time and check only study as category of mobility as she is not looking for any type of internship at the moment. She then select his home country and begin to browse among the different flows in the top-right map that shows the Italian students that go abroad. Thanks to the difference in thickness of the lines she immediately notices that the biggest flow is the one of Italian students going to Spain, and the following most common destinations are France, Germany and the United Kingdom: she is not surprised by this result as she expected these to be the main destination from all over Europe. Moreover, she can move around the map and get general summaries about the main destinations and the variety of other possibility. By hovering over different lines she gets the tooltip with the bar chart that show the main field of study that characterize each flow: in this way she discovers that in most of the cases *Science, Mathematics and Computing*, that comprises his course of *Computer science*, is not on the top of the field rank and just a few of the students of each flow belongs to that field. Nonetheless, she discovers that is the fourth major field of study in the United Kingdom and that among all destination there is an approximately even distribution of the two genders. We report the details of the analysis in Figure9

Completed the exploration of the *Destination dashboard*, she now starts to look at the *Home dashboard* to have a more in depth point of view regarding his university of Bari: she starts by lookup and select Italy in the heatmap of Europe and subsequently select Bari in the map of Italy, while she cannot fail to notice the diversity in the absolute number of outgoing students from different cities in Italy. Finally she clicks on her university and the application moves automatically to a detailed view of the university to proceed in a more in-depth analysis. in the lower part of the page she firstly notice that the majority of students belong to *Social Sciences, Business and Law*, *Health and Welfare* and *Humanities and Arts* and, on the other side, that students from *Science, Mathematics and Computing* on average gain less ECTS (21.98) that the other main field. Thus, she filter on her course in Computer science and this gives her the specific subset where she belongs and the data of similar students that have already been abroad. She discover that for her course the average

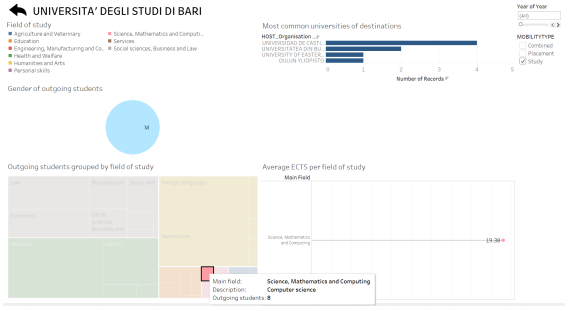

Fig. 10. Dashboard of outgoing students from *Università degli studi di Bari* after having selected *Computer science* as course of study

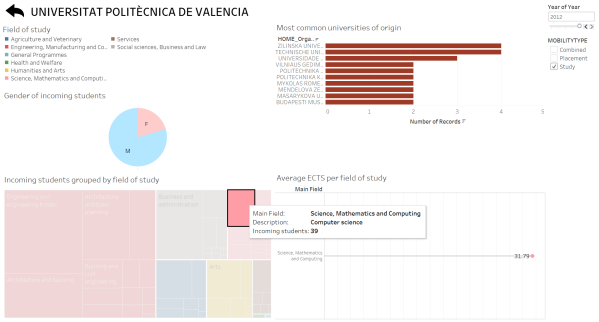

Fig. 11. Detail for incoming student in *Universitat politècnica de Valencia*

number of ECTS gains is even lower (19.38) than the average for the main field and that all (8) previous students that have been abroad were males. Having said that, she notices that half of the students have been to Universidad de Castilla-La mancha so she aim to discover something more about that university. The final dashboard is shown in Figure 10.

She then change dashboard and moves to the *Host dashboard* where she want to find out which universities are more suitable for her preferences. The map on the top right corner presenting the absolute number of incoming students per country confirm her expectations on the most visited countries as previously expressed. Given the previous result of Universidad De Castilla-La mancha of Toledo as the most frequented by her colleagues she select the Spain and browse over different city. By filtering on only one year, for example on the most recent as it is 2012, she get a general idea on the number of students going to each city every year: Toledo seems quite small and she opt for Valencia, in particular she starts by looking at *Univesitat politècnica de Valencia*. In that university the majority of students are related to *Engineering, Manufacturing and Construction*, but *Science, Mathematics and Computing* also has an important role. Moreover, the average ECTS gains by this students is 27.59 and when she selects his course of study the average value of her colleagues is even higher(31,79). She notice that in this university, for her course, there is less than a quarter of females students, but by selecting female students in the pie chart she notices that, on average, they gain an even higher number of ECTS.

As a final task, she wants to compare Italy with Spain as she wants to understand whether there are some discrepancies, so she moves to *Comparison dashboard*. She first select Italy and Spain in the map (by tilting or using crtl) and then she

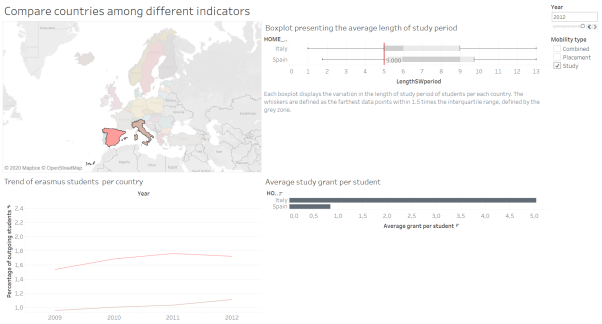

Fig. 12. Dashboard that compare characteristics for Italy with Spain

filters out both Combined and Placement type of mobility so she can have a better idea of the students characteristics. She can now see that both the countries experienced an increase in the number of students during the four year that are considered and that in Spain the percentage of students that participate to the Erasmus program is always higher then 1,5% while for Italy the same data is around 1%. Also, the mobility of both Italian and Spanish student tend to last more compared to the average of all the countries (5.93 months): Italians have an average just below 7 months and Spanish around 8 months of mobility. In the bottom-right corner, instead, the user realizes that Spanish people get more money on average for the mobility than Italians: by hovering on the bars of the two country she notices that Spain distributes the grand in a more continuous way, while in Italy there are a lot of gaps. The dashboard that has been described can be seen in Figure12.

4.2 A journalist in the Netherlands

The second use case we define is the one of a journalist from Amsterdam that want to write a report about the Erasmus program and, moreover, want to investigate and verify the general idea that the number of students that participate to that program is increasing.

To have a general overview of the program regarding the whole Europe he starts from the *Comparison dashboard* and filter for the oldest year (2012) and the three mobility types at the same time. At a glance, in the *Trend of erasmus student per country* section, he notices that Luxembourg and Liechtenstein have percentages of outgoing students that stand from the rest of the country: around 8% and 4% respectively during the four year, while for the rest of the countries the percentages are all under the 2%. To have a more complete view of the data he filter out Luxembourg and Liechtenstein and obtain a more precise graph where it is possible to notice more subtle differences between the remaining countries. For example, thanks to this in-depth view, it is possible to see that for the first two year Spain was the country with the higher percentage of students (1.94% in 2010), but from 2011 on Latvia overtake this role (2.11%). On the lower part of the graph he notices that Turkey has the lowest percentage with a steady value around 0.3%. By hovering on the lines representing each country he can have an even more precise view with the areal graph representing the absolute number of students grouped by field of study. Indeed, most of the countries seem to have a positive trend and he can conclude that overall the number of students that participate to the program has increased. After having filter in Luxembourg and Liechtenstein he moves to the right side of the dashboard and specifically to the *Boxplot*

presenting the average length of study period: the red line show the average length of the study period among all countries. He can easily conclude that most of the students spend only one semester abroad. Moreover, he notice that lots of countries have even lower values and only Spain, France, Italy, United Kingdom and Ireland have values that are over the average of the entire Europe.

Proceeding in the comparison he looks at the *Average study grant per student* tab that uses a bar chart to summarizes the average study grant received by the students in each country. Immediately he notices that Liechtenstein has the maximum value per student in 2009 and 2012, but have 0€ for 2010 and 2011. Indeed he expect to have some missing value related to the grant received by these students in 2010 ad 2011.

After the general overview, he select Netherlands on the map to have a detailed report of his country: the percentage is growing till 2011 when it reach 1.37%, but in 2012 it decreases. Nonetheless, the absolute number of student that appears in the tooltip shows a positive trend with a clearly increasing number of students in *Social science, Business and Law*. The average length of period varies between 3 and 6 months and is on average less than the mean for the Europe. The average grant is just above 1000€ per student with an even distribution between males and females. Moreover, there were a high number of people that gain a grant of 1000€ and 1200€.

He now moves to the *Host dashboard* to have an idea of the students that came in the Netherlands to enjoy their opportunity. Using the heat map on the top he understand that the Netherlands have not a high percentage of erasus students, but they still count for the 0.5% in the whole population of students. After selecting his country, he start to browse the map of the Netherlands where the main cities appears to have the higher number of foreign students as he would expect.

5 Discussion

Our methodologies are based on the scientific literature presented and suggested during the Visualization course. In particular, we relied especially on [8] to clearly define our tasks and [3] to critically analyze and implement them. Each task has been define according to the taxonomy presented in the first paper, thus allowing us to have a clear goal and ease our visualization design choices. Lastly, we could have a more structured approach for the implementation of the various dashboards: defining a hierarchy on WHY, HOW and WHAT made the work more structured. For the implementation of the origin-destination flow map, as discussed in 3.1, we strongly relied on the work of B. Jenny et al. [10] as this helps us to create a simply and easy-to-read visualization and clearly improve our first solution where we used straight lines. It is worth to say that, due to disparity in the dataset among different `mobility types`, we had to combine `COUNTRYOFWORKPLACEMENT` and `COUNTRYOFHOSTINSTITUTION` to be able to present different flows in the same visualization.

Regarding the realization of the two symmetrical dashboards concerning home and host institutions described in 3.2, we started by applying some principles and ideas given by J. Heer et al. [9] and subsequently correcting our initial choices influenced by the reflections and observations pointed out in the work of D. Borland and R.M. Taylor II [2]. Indeed, for the heat-maps presenting the incoming and outgoing students in each country, we changed the absolute values with relative ones in order to show percentage as it is not misleading. Nevertheless, next to the heat-maps, we chose to represent the absolute number in a graduated symbol map where the number of students is encoded with the size of the circle for each country. Our choice is justified because, even if the circular areas might not be easy to read and compare, this solution gives us an intuitive idea of discrepancies within Europe. For the same reason, we decided to use this approach for both cities and universities with a graduated symbol map and a packed bubbles chart respectively. In particular, in order to improve the readability for the universities, we encode the absolute number of students coming from/going to each university in both the color and the size of the circles. We have considered a more classical approach using bar-charts as well as a tree-maps, but since we were not interested in the actual analytical comparisons of different universities, we were satisfied with a general impression of the size of each city/university.

Regarding the in-depth view of a university we combined a number of various technique to both add values to the analysis and help the user to distinguish between different results. In particular we have to carefully choose to color of different classifier (`GENDER` and `MAINFIELD`) to not incur in overlaps of color mapping that would have created misleading contents: indeed, taking into account the considerations of M. Stone [12], we choose pastel colours for the mapping of genders and more vivid ones to distinguish between main-fields. For the gender we use a pie chart because it only has two classes that entirely cover the set of possibilities offered. Alternatively we considered the use of a stacked bar chart, but in this case we have just one main group and do not have to compare with other universities, so the result is both complete and easy to understand. For the field of study of the students we finally opted for a tree map. In this case we considered other alternatives layout presented in [9] as icicle tree, sunburst and nested circles. Most of them were inappropriate for the use of the space and the readability. We also tried to use stacked bar chart but, even if in this case the differences between main fields were more clear, it was harder to deep into the level of study program. We use a bar chart to represent the most common universities of destination as it clearly show the difference between destinations and it is easy to retrieve the absolute number of students. For the average ECTS grouped by field of study the first choice was a line graph with the time coordinates to express the progress during the considered period. We found it was not a good choice and decided to use a filter for years instead. Next we tried to use a bar chart, but because the values represent averages of absolute values and not the number of student achieving a particular number of ECTS we opted for the lollipop chart as it appear in the final version.

For the dashboard used to compare different countries we use a map for the selection of the country as it is the simplest and clearer way to present the data. With the same approach of simplicity we use a line graph to represent the percentages of student participating in the program each year and a bar chart for the average grant offered per each students. In both cases, we report other graphs when a country is selected and this help to understand the data at a deeper level of detail.

6 Conclusion

Build a visualization tool to analyze a complex dataset such as the one we took into consideration requires intensive work in both the preparation phase of the data and the actual implementation of the tool itself. We realized how the definition of tasks a priori is essential to move through all the possible visualization that can be used with the available data. Once the tasks have been defined, each of them must be tackled in an incremental way since, sometimes, there is a need to integrate the dataset somehow with transformed features or, as in our case, there is a need to find a proper mapping. Our visualization has been able to answer

multiple questions that might arise by thinking about the Erasmus program, such as which the city is preferred the most in a certain country or in general for a precise year and mobility type. We believe that our tool can be extended to more years so, besides the already implemented tasks, other techniques to explore temporal trends in detail can be exploited. To conclude, we consider our final work a useful and interactive tool to investigate data about the Erasmus program, an organization that plays an essential role in blurring the boundaries between different cultures, leading the next generations towards an international and genuinely open society.

WORK DIVISION

In the team we had a fair and equal division of the work, possible mainly because we are both interested in the power of visualization as a tool to communicate fact and data. We separately read the visualization papers mention during lectures to have the necessary knowledge to first define the tasks and later develop the dashboards. Following a brainstorming session, we decided on which kind of topic we wanted to focus on and therefore we ended up choosing the dataset about Erasmus found in the EU Open Data Portal [1] with the aim to have a high level of details as well as a time-series data to analyze possible trend during time. After having defined the dataset, each of us came up with a set of possible tasks and together we selected the most interesting ones. Thereafter, each of us started working on a different tasks and, after multiple feedback sessions, we combined them together. In the initial period, we met once a week to discuss the progresses and possible difficulties and, in the final period, we met more frequently to combine all the tasks together and hone the entire visualization. In order to write the report, instead, we decided to work on different sections so we could optimize the work. After the completion of a task, we had to review it together, discuss it and make it conclusive.

REFERENCES

[1] Eu open data portal. https://data.europa.eu/euodp/en/data/.

[2] D. Borland and R. M. T. II. Rainbow color map (still) considered harmful. *Visualization Viewpoints*, 2007.

[3] M. Brehmer and I. Tamara Munzner, Member. A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics*, 2013.

[4] S. Directorate-General for Education, Youth and Culture. https://data.europa.eu/euodp/en/data/dataset/erasmus-mobility-statistics-2008-09, 2018.

[5] S. Directorate-General for Education, Youth and Culture. https://data.europa.eu/euodp/en/data/dataset/erasmus-facts-figures-trends-2009-2010", 2018.

[6] S. Directorate-General for Education, Youth and Culture. https://data.europa.eu/euodp/en/data/dataset/erasmus-facts-figures-trends-2010-2011, 2018.

[7] S. Directorate-General for Education, Youth and Culture. https://data.europa.eu/euodp/en/data/dataset/erasmus-mobility-statistics-2011-12, 2018.

[8] M. H. Hans-J¨org Schulz, Thomas Nocke and H. Schumann. A design space of visualization tasks. *IEEE transactions on visualization and computer graphics*, 2013.

[9] M. B. Jeffrey Heer and S. U. Vadim Ogievetsky. A tour through the visualization zoo. *Acmqueue*, 2010.

[10] B. Jenny, D. Stephen, I. Muehlenhaus, B. Marston, R. Sharma, E. Zhang, and H. Jenny. Design principles for origin-destination flow maps. *Cartography and Geographic Information Science*, 12 2016.

[11] P. Riehmann, M. Hanfler, and B. Froehlich. Interactive sankey diagrams. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS*, pages 233 – 240, 11 2005.

[12] M. Stone. Choosing colors for data visualization. http://www.b-eye-network.com/newsletters/ben/2235, 2006.

[13] B. Team. Good chart, bad chart. https://medium.com/biovinci/good-chart-bad-chart-b6a6725f8795.